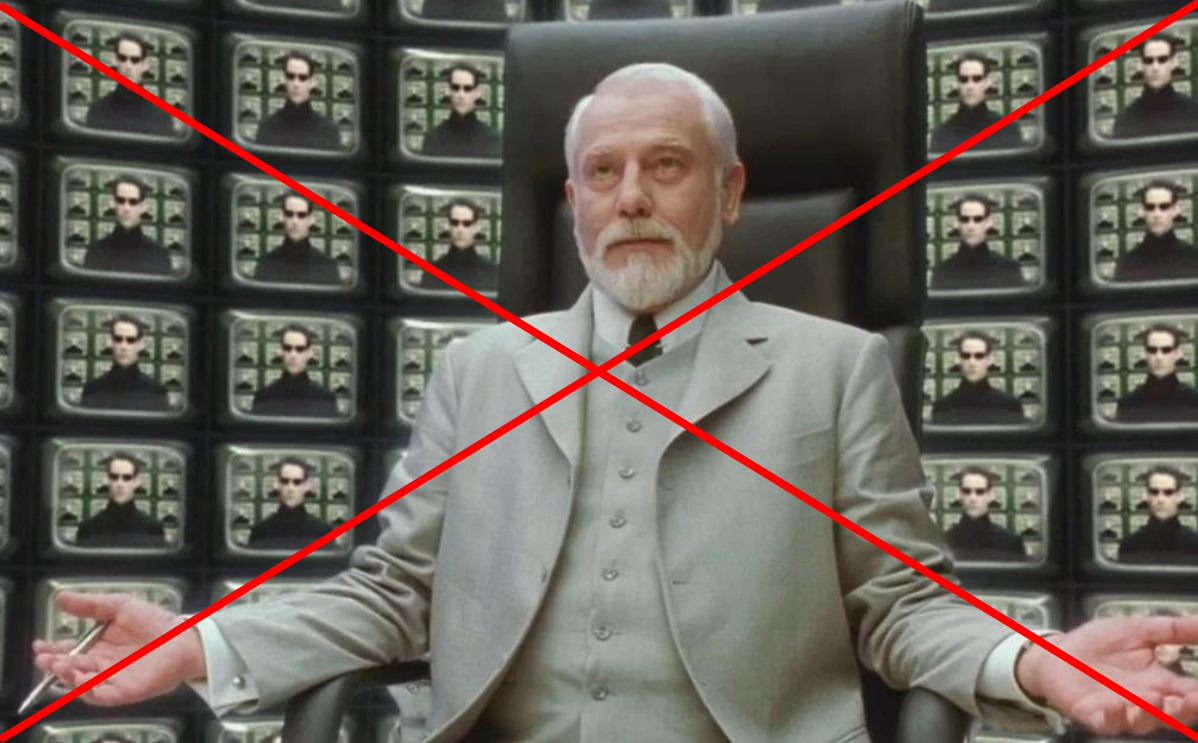


# НЕЙРОННЫЕ СЕТИ В ОБРАБОТКЕ ЕСТЕСТВЕННЫХ ЯЗЫКОВ: ОБЗОР НОВОСТЕЙ В ПРИМЕНЕНИИ К БИБЛИОТЕЧНЫМ ПРОЦЕССАМ

ОБЗОР ВОЗМОЖНОСТЕЙ

декабрь 2022



# ИИ – ЭТО НЕ ИНТЕЛЛЕКТ


**ИИ – это набор математических алгоритмов, умеющих отыскивать закономерности в данных гораздо лучше и быстрее, чем это может сделать человек.**  
Найденные закономерности затем используются.



ВОЮТ НЕ МАШИНЫ, ВОЮТ ЛЮДИ!!!

# ОТ ПОИСКА ЗАКОНОМЕРНОСТЕЙ - К ПОНИМАНИЮ ТЕКСТА

погода в ✕ Найти

погода в ижевске  -8 °C ↖

погода в ижевске на 10 дней ↖

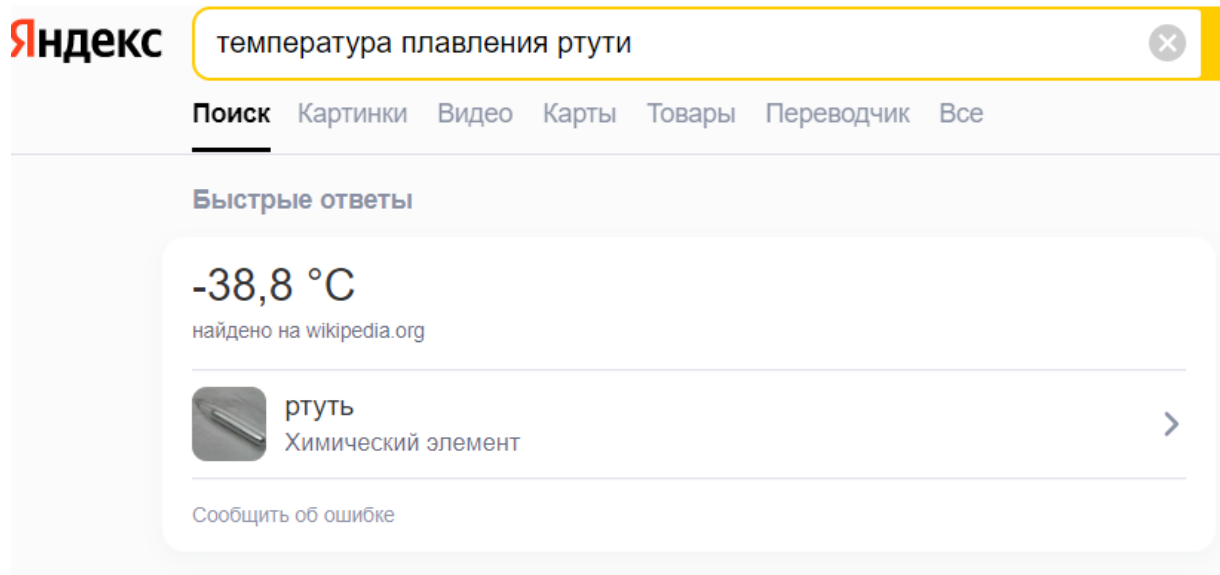
погода в ижевске на 3 дня ↖

погода в ижевске на неделю ↖

погода в ижевске на месяц ↖

- Распознавание структуры текста
- Боты - собеседники и помощники, FAQ и QA-машины
- Выявление перефразирований
- Классификация документов по различным критериям
- Ранжирование и поиск
- Выделение дат и именованных сущностей
- Построение графа отношений между именованными сущностями
- Представление документа в виде графа знаний

# ОТ ПОИСКА ЗАКОНОМЕРНОСТЕЙ - К ПОНИМАНИЮ ТЕКСТА



- Распознавание структуры текста
- Боты - собеседники и помощники, FAQ и QA-машины
- Выявление перефразирований
- Классификация документов по различным критериям
- Ранжирование и поиск
- Выделение дат и именованных сущностей
- Построение графа отношений между именованными сущностями
- Представление документа в виде графа знаний

# ВЫЯВЛЕНИЕ ПЕРЕФРАЗИРОВАНИЙ

Модель семантического следования русского языка обучена на датасете TERRa (Textual Entailment Recognition for Russian). Модель представляет собой задачу бинарной классификации является ли значение одного текста вытекающим (может быть выведено) из другого текста в данной паре.

## Введите утверждение

Предприниматели продолжают устанавливать незаконные постройки, однако власти города на примере Сенной площади уже доказали, что лучше не спорить и убрать свое добро самостоятельно.

## Введите гипотезу

Предприниматели иногда возводят строения незаконно.

Спросить

## Результаты

entailment

Предприниматели иногда возводят строения незаконно.

Предприниматели продолжают устанавливать незаконные постройки, однако власти города на примере Сенной площади уже доказали, что лучше не спорить и убрать свое добро самостоятельно.

## Примеры

Автор поста написал про канализацию.

Предприниматели иногда возводят строения н...

Кризис обязательно рассмотреть как признак п...





# ПОИСК И РАНЖИРОВАНИЕ: TF-IDF VS НЕЙРОННЫЕ СЕТИ

## TF-IDF (частотность слов)

- Лучше работает с длинными текстами
- Быстрый
- Входит в состав стандартных поисковых программных продуктов (ElasticSearch, Solr)

## Нейронные сети (на основе трансформеров)

- Работает только с короткими фрагментами (1-2 абзаца)
- Медленный
- Понимает перефразирования
- Используется в качестве средства, уточняющего ранжирование

5 лучших нарядов для учителей

Читайте на > странице 14.

# КОМСОМЛЬСКАЯ ПРАВДА

Волгоград

Радио КП 96.5 FM



Пятница - суббота, 3 - 4 октября

Газета нашего города

№ 112 (28290) 2014 г.

Регистрационная цена 12 руб.

Олег Царев в Волгограде:

## НАМ НУЖНЫ ДОБРОВОЛЬЦЫ ОТСТРАИВАТЬ ДОНБАСС



Подробности по ссылке >

### Спокойно, мамы, капитал будет вашим

Отмена программы материнского капитала не отменит ни льготы, ни выплаты. Об этом сообщила пресс-секретарь главы Кабинета Наталья Тиманова.



По ее словам, программа будет действовать, как и запланировано, до 2016 года. Так она сообщила в ходе встречи с журналистами, что Минэкономразвития якобы предлагает заменить предоставление маткапитала.

### Дмитрий Шендерович привез сына к Жанне Фриске

Жанна Фриске улетела в Лигу в Негай, где будет проходить дальнейшее обследование у местных врачей. По словам Дмитрия Шендеровича, он привез сына к Жанне Фриске.



Жанна Фриске улетела в Лигу в Негай, где будет проходить дальнейшее обследование у местных врачей. По словам Дмитрия Шендеровича, он привез сына к Жанне Фриске.

Телеведущий Александр ВАСИЛЬЕВ учил депутатов Госдумы выгадать стилисты: Новая русская мода - с утра кокошник, вечером бриллианты

Елена ЧЕРНОВА, Юлия ХОМЯТОВА

«КП» публикует самые интересные выдержки из лекции историка моды.

Такого в Госдуме еще не видели: перед парламентом во вторник прозвучала лекция историка моды и телеведущий Александр Васильев в стальных очках и шарфе. Хотите каждое утро уметь с толковым одесским простым жаргоном - пора выучить «Моду и историю» в слухом формате, то есть аудиоподкасте.

Продолжение читайте по ссылке >

# ВЫДЕЛЕНИЕ СТАТЕЙ ИЗ ПОЛОСЫ

Возможности ИИ позволяют автоматизировать выделение статей из распознанного текста полос издания: таким образом, есть возможность создавать полнотекстовые библиотечные записи не по полосам, а по статьям, что гораздо удобнее для использования.

ПОЧТИ 100% ТОЧНОСТЬ



Честь Тюменской области на чемпионате защищают четверо учащихся из Западно-Сибирского государственного колледжа и регионального лесотехникума, в том числе представительница прекрасного пола. Еще один претендент на победу прибыл на соревнование из Пермского края.

Задание участникам турнира досталось красивое и необычное: собрать деревянную беседку с романтическим названием «Козырек для поцелуев».

Беседка состоит из трех частей, каждая из которых собирается отдельно. В этом году конкурсное задание усложнилось тем, что плотникам не предоставляют готовые чертежи. Для третьего модуля они самостоятельно, за три месяца перед профессиональным турниром, должны были посчитать количество нужных деталей и расположить их на бумаге в четкую понятную схему.

Конкурсанты, соревнующиеся на World Skills Russia, получают заманчивые предложения от работодателей. Региональный чемпионат профмастерства практически гарантирует талантливым мастерам по прическам трудоустройство. Об этом рассказала главный эксперт.

– Мы следим за тем, как ребята выполняют техническую стрижку с креативной укладкой, сложную прическу, а также женскую салонную и классическую мужскую стрижку с укладкой, – пояснила главный эксперт Светлана Юровская. – Все конкурсанты – это парикмахеры-универсалы. Таких специалистов работодатели забирают с конкурса охотно!

Добавим, что в региональном этапе чемпионата World Skills Russia участвует шесть конкурсантов. Трое из них представляют областной центр, двое приехали из Тобольска и один участник – ишимский.

# ВЫДЕЛЕНИЕ ИМЕНОВАННЫХ СУЩНОСТЕЙ И СВЯЗЕЙ МЕЖДУ НИМИ

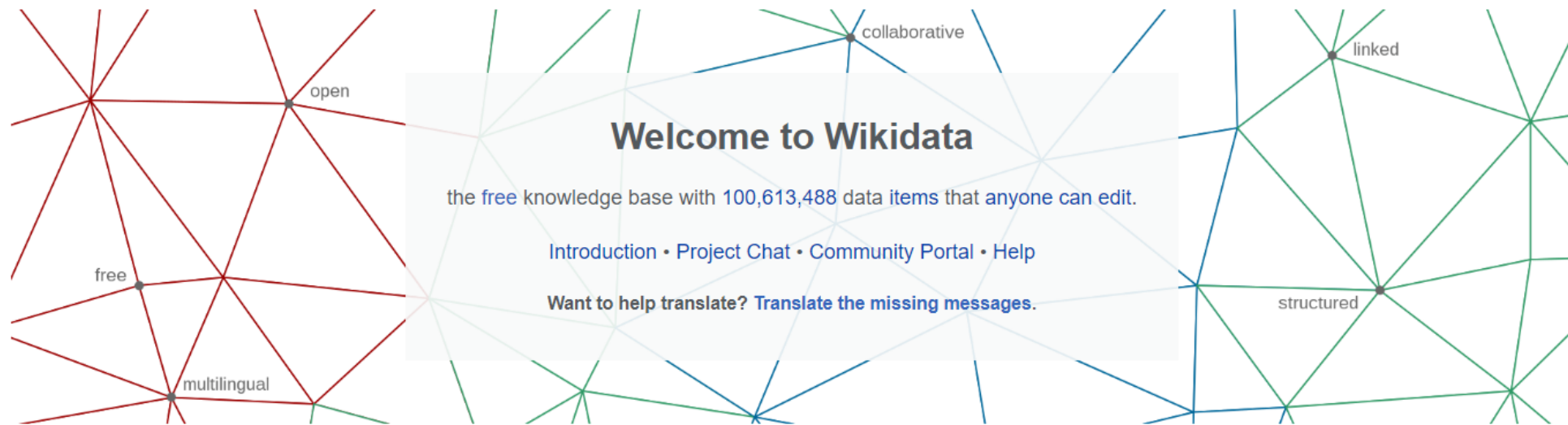
«Читая» распознанный текст, ИИ может выделить из текста имена собственные:

географические названия, имена и фамилии, названия событий итд.

Между выделенными сущностями выстраиваются связи в формате Wikidata.



# ПРОЕКТ WIKIDATA: 100 613 488 СУЩНОСТЕЙ НА ВЧЕРА



## Welcome!

Wikidata is a free and open knowledge base that can be read and edited by both humans and machines. Wikidata acts as central storage for the **structured data** of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wiktionary, Wikisource, and others.

Wikidata also provides support to many other sites and services beyond just Wikimedia projects! The content of Wikidata is [available under a free license](#), [exported using standard formats](#), and can be [interlinked to other open data sets](#) on the linked data web.

## Get involved

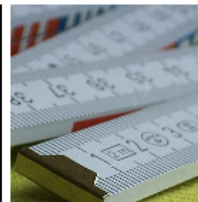
For a complete starters' guide, visit the [community portal](#).

## Learn about data

New to the wonderful world of data? [Develop and improve your data literacy through content](#) designed to get you up to speed and feeling comfortable with the fundamentals in no time.



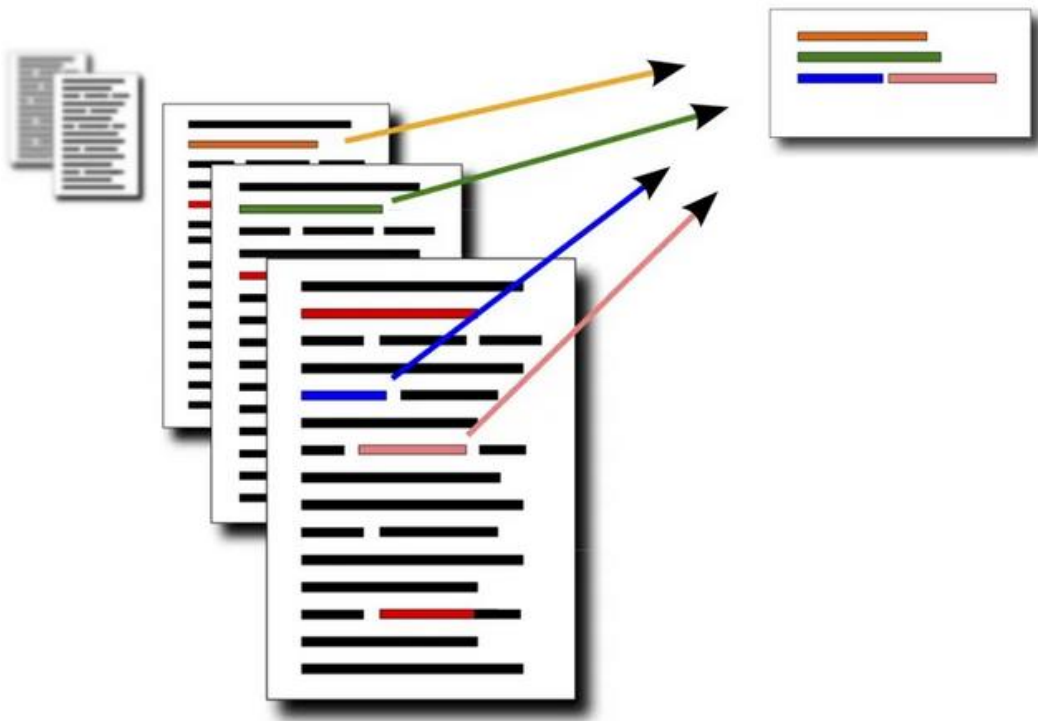
Item: *Earth* (Q2)



Property: *highest point* (P610)



custom value:  
*Mount Everest*  
(Q513)



# РЕФЕРИРОВАНИЕ ТЕКСТА СТАТЬИ

**«Прочтя» статью, ИИ может сформировать краткое изложение её текста, «реферат», который затем может использоваться при формировании библиографической записи:**  
этот процесс называется суммаризацией.

СПАСИБО



Лютецкий Владимир



+7 (903) 640 2815



[voljgm@gmail.com](mailto:voljgm@gmail.com)